#5

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of

Kenji YAMANISHI, et al.

Appln. No.: 09/805,114

Confirmation No.: 1503

Filed: March 14, 2001

For: QUESTIONNAIRE ANALYSIS SYSTEM

Docket No: Q63084

Group Art Unit: 2121

Examiner: Joseph P. HIRL

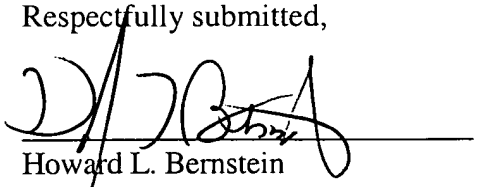
SUBMISSION OF VERIFIED ENGLISH TRANSLATION OF PRIORITY DOCUMENT

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Submitted herewith is a verified English translation of the priority document (Japanese Patent Application No. 2000-071657) on which the claim to priority was made under 35 U.S.C. § 119. The Examiner is respectfully requested to acknowledge receipt of said verified translation.

Respectfully submitted,



Howard L. Bernstein
Registration No. 25,665

SUGHRUE MION, PLLC
Telephone: (202) 293-7060
Facsimile: (202) 293-7860

WASHINGTON OFFICE
23373
CUSTOMER NUMBER

Date: December 23, 2003

In the matter of

US Patent Application No.09/805,114

DECLARATION

I, KAZUO HAYASHI, c/o YAMASHITA & ASSOCIATES of Toranomon 40th MT Bldg., 13-1, Toranomon 5-chome, Minato-ku, Tokyo 105-0001, Japan, do sincerely declare that I well understand the Japanese language and the English language and that the attached English translation of a certified copy of Japanese Patent Application No.071657/2000 is a true, correct and faithful translation to the best of my knowledge and belief from the Japanese language into the English language.

December 16, 2003

Kazuo Hayashi

Kazuo Hayashi

(Translator)

[Name of Document] APPLICATION FOR PATENT

[Reference Number] 33509710

[Addressee] Hon. Commissioner of Patent Office

[IPC] G06F 17/00

[Inventor]

[Domicile or Residence] c/o NEC corporation,
7-1, Shiba 5-chome, Minato-ku, Tokyo

[Name] YAMANISHI Kenji

[Inventor]

[Domicile or Residence] c/o NEC corporation,
7-1, Shiba 5-chome, Minato-ku, Tokyo

[Name] RI Ko

[Applicant for Patent]

[Identification Number] 000004237

[Name] NEC Corporation

[Agent]

[Identification Number] 100088890

[Patent Attorney]

[Name] KAWAHARA Junichi

[Indication of Fee]

[Deposit Account Number] 009690

[Amount of Payment] 21,000

[List of Attached Documents]

[Name of Document] Specification 1

[Name of Document] Drawings 1

[Name of Document] Abstract 1

[Number of General Power of Attorney] 9001717

[Proof] Necessary

[Name of the Document] Specification

[Title of the Invention] INQUIRY REPLY ANALYSIS SYSTEM

[Scope of Claim for a Patent]

[Claim 1] An inquiry reply analysis system comprising:

means for entering an inquiry reply statement including free reply description in natural language,

a network for transmitting the inquiry reply statement,

a database for accumulating the transmitted inquiry reply statements, and

a text classification engine for reading out the inquiry reply statements from the database and learning a rule for classifying the inquiry reply statements.

[Claim 2] An inquiry reply analysis system comprising:

means for entering an inquiry reply statement including free reply description in natural language,

a database for accumulating the transmitted inquiry reply statements, and

a text classification engine for reading out the inquiry reply statements from the database and learning a rule for classifying the inquiry reply statements.

[Claim 3] An inquiry reply analysis system comprising:

means for entering an inquiry reply statement including free reply description in natural language,

a network for transmitting the inquiry reply statement,

a database for accumulating the transmitted inquiry reply statements,

a text classification engine for reading out the inquiry reply statements from the database and learning a rule for classifying the inquiry reply statements, and

means for distributing the rule through the network according to a

request from a claimant.

[Claim 4] An inquiry reply analysis system according to claim 1, 2, or 3,

wherein said text classification engine includes morpheme analysis means for analyzing morphemes in all sentences in the inquiry reply statements accumulated in the database, category-text designating means for designating the category and text, attribute selecting means for selecting attributes in plural inquiry reply statements being read in from the database, rule learning means for learning the rule for expressing the correspondence of the text and category on the basis of the words selected by attributes by the attribute selecting means, and rule output means for issuing the rule learned by the rule learning means.

[Claim 5] An inquiry reply analysis system according to claim 4,

wherein said attribute selecting means executes a procedure of computing the difference $\Delta SC(\omega)$ between the stochastic complexity of a test set without consideration of appearance of word and the stochastic complexity of a text set with consideration thereof, in each word appearing in the text, and a procedure of selecting as attribute when the difference $\Delta SC(\omega)$ is lager than the threshold τ.

[Claim 6] An inquiry reply analysis system according to claim 4,

wherein said rule learning means executes a procedure of forming the text set by replacing with an expression of (d_1, c_1), (d_2, c_2), ..., (d_m, c_m) [where each d_i is a multi-valued discrete vector $d_i = (\omega_{i1}, \omega_{i2}, ..., \omega_{in})$ $(i = 1, ..., m)$, ω_{ij} is 1 when word obtained by attribute selection ω_j $(j = 1, ..., n)$ appears in the i-th text, or 0 otherwise, c_i expresses the value (label) of the category according to the i-th text and each c_i is 1 when belonging to a specific category, or 0 otherwise, and m is the number of texts], a procedure of growth process by selecting the rules of if-then-else format and sequentially adding to the stochastic decision list by employing the information quantity standard such as the extended stochastic complexity minimum principle or stochastic

complexity minimizing principle, and a procedure of removing the rules one by one from the last one of the stochastic decision list, and clipping continuously until none should be removed from the viewpoint of the extended stochastic complexity minimum principle.

[Claim 7] A recording medium recording a program for making a computer function as morpheme analysis means for analyzing morphemes in all sentences in the inquiry reply statements accumulated in the database, category-text designating means for designating the category and text in the text classification engine, attribute selecting means for selecting attributes in plural inquiry reply statements being read in from the database, rule learning means for learning the rule for expressing the correspondence of the text and category on the basis of the words selected by attributes by the attribute selecting means, and rule output means for issuing the rule learned by the rule learning means.

[Detailed Description of the Invention]

[0001]

[Technical Field of the Invention]

The present invention relates to an inquiry reply analysis system, and particularly to an inquiry reply analysis system using a text automatic classification technology, a natural language processing technology, and a network utilization technology.

[0002]

[Prior Art]

The operation for extracting general features and tendency from inquiry reply statements including free reply description in natural language obtained through the network such as the Internet has been conventionally done almost manually. Text mining tools such as DE-FACTO developed by Dentsu (published in leaflet), Keyword Associator of Fujitsu (I. Watanabe: Divergent thought support system "Keyword Associator" 2nd edition,

research group paper of 15th Meeting of System Engineering Group of Society of Measurement and Automatic Control of Japan, July 1994), and "HIPS" (Watanabe, Miki, Nitta, Sugiyama: Hybrid thought support system HIPS, research group paper of 15th Meeting of System Engineering Group of Society of Measurement and Automatic Control of Japan, January 1995) were used for extracting the relationship of words from the text information. However, these tools could not express the features of inquiry reply statements in a format of rule.

[0003]

So far, nothing has been known about the system or service for collecting and analyzing inquiry reply statements including free reply description in natural language automatically through the network such as the Internet, and distributing the analysis results, if necessary, to the claimant. For example, in JP H11-066036A or JP H11-143856A, the technology for inquiring through the network and accumulating the replies in the database is disclosed, but features of inquiry reply statements are not extracted in a format of rule.

[0004]

[Problems that the Invention is to Solve]

In the conventional manual inquiry reply analysis mentioned above, if the number of inquiry replies is tremendous, the operation is too inefficient.

[0005]

In text mining tools such as DE-FACTO and HIPS, features of inquiry replies cannot be extracted in a format of rule, it was not sufficient from the viewpoint of presentation of compact and easy knowledge.

[0006]

The text classification tool used in retrieval is a useful tool for analysis of inquiry replies, but it was not used before in analysis of inquiry replies including free reply description in natural language.

[0007]

It is a first object of the invention to present an inquiry reply analysis system capable of automatically presenting knowledge in a compact and easy rule from inquiry reply statements including free reply description in natural language by using a text classification engine.

[0008]

It is a second object of the invention, in addition to the above, to present an inquiry reply analysis system for distributing analysis results to the claimant by automatically extracting the knowledge in the rule format from the inquiry reply statements collected through the network.

[0009]

[Means for Solving the Problems]

An inquiry reply analysis system of the invention comprises means for entering an inquiry reply statement including free reply description in natural language, a network for transmitting the inquiry reply statement, a database for accumulating the transmitted inquiry reply statements, and a text classification engine for reading out the inquiry reply statements from the database and learning a rule for classifying the inquiry reply statements.

[0010]

Further, an inquiry reply analysis system of the invention comprises means for entering an inquiry reply statement including free reply description in natural language, a database for accumulating the transmitted inquiry reply statements, and a text classification engine for reading out the inquiry reply statements from the database and learning a rule for classifying the inquiry reply statements.

[0011]

Moreover, an inquiry reply analysis system of the invention comprises means for entering an inquiry reply statement including free reply description in natural language, a network for transmitting the inquiry

reply statement, a database for accumulating the transmitted inquiry reply statements, a text classification engine for reading out the inquiry reply statements from the database and learning a rule for classifying the inquiry reply statements, and means for distributing the rule through the network according to a request from a claimant.

[0012]

[Mode for Carrying Out the Invention]

Embodiments of the invention are described in detail below while referring to the accompanying drawings.

[0013]

(1) First Embodiment

Fig. 1 is a block diagram showing a configuration of an inquiry reply analysis system according to a first embodiment of the invention. The inquiry reply analysis system of the embodiment mainly comprises respondent computers 111 to 11N (N being a positive integer), a network 12, a database 13, and a text classification engine 14.

[0014]

The respondent computers 111 to 11N are computers, portable information terminals, cellular phones, and other devices having transmission function of message, mail and the like, which are connected to the network 12.

[0015]

The network 12 includes various networks, whether wired or wireless, such as public networks, exclusive networks, or LAN (local area network).

[0016]

The database 13 is connected to the network 12, and inquiry reply statements from plural respondents transmitted from the respondent computers 111 to 11N through the network 12 are accumulated herein.

[0017]

The text classification engine 14 reads out plural inquiry reply statements from the database 13, extracts a rule for classifying the inquiry reply statements, and issues the rule to the claimant. The text classification engine 14 includes morpheme analysis means 15 for analyzing morphemes in all sentences in the inquiry reply statements accumulated in the database 13, category-text designating means 16 for designating the category and text in the text classification engine 14, attribute selecting means 17 for selecting attributes in plural inquiry reply statements being read in from the database 13, rule learning means 18 for learning the rule for expressing the correspondence of text and category on the basis of the words selected by attributes by the attribute selecting means 17, and rule output means 19 for issuing the rule.

[0018]

The text classification engine 14 is an engine for learning the corresponding relation of the category and text as a classification rule, and, for example, an engine proposed by Li and Yamanishi can be used (H. Li and K. Yamanishi: Text Classification Using ESC-based Stochastic Decision Lists, Proceedings of 1999 International Conference on Information & Knowledge Management, pp. 122-130, 1999). This text classification engine basically conforms to the system of "Forming method and apparatus of decision list" disclosed in Japanese Patent No. 2581196.

[0019]

Fig. 2 shows a composition of a set of inquiry reply statements accumulated in the database 13. Each row expresses an inquiry item, and each line shows the reply statement of one person.

[0020]

Referring to Fig. 3, processing of the text classification engine 14 comprises morpheme analysis step 31, designating step 32 of text and category, attribute selecting step 33, rule learning step 34, and rule output

step 35.

[0021]

Referring to Fig. 4, a more specific processing of attribute selecting step 33 includes $\Delta SC(\omega)$ computing step 41, and attribute selecting step 42.

[0022]

Referring to Fig. 5, a more specific processing of rule learning step 34 includes data forming step 51, growth processing step 52, and clip processing step 53.

[0023]

Fig. 6 is a diagram showing an example of rule format knowledge (stochastic decision list) as a result of analysis by the text classification engine 14.

[0024]

Fig. 7 is a diagram showing other example of rule format knowledge (stochastic decision list) as a result of analysis by the text classification engine 14.

[0025]

In the inquiry reply analysis system of the first embodiment having such configuration, the operation is explained below.

[0026]

When inquiry respondents send inquiry reply statements from the respondent computers 111 to 11N, the inquiry reply statements are communicated in the database 13 through the network 12. Suppose the number of respondents to be N. At this time, the inquiry reply statements may include free reply description in natural language.

[0027]

The text classification engine 14, first by the morpheme analysis means 15, analyzes morphemes in all sentences of inquiry reply statements accumulated in the database 13 (step 31).

[0028]

Next, by the category-text designating means 16, the text classification engine 14 causes the operator to designate the category and text in the inquiry reply statements (step 32). Herein, designation of category is to classify by paying attention to the replies in one row. For example, it is the category designation that, relating to the first row in Fig. 2, the replies are classified into "company A", and "other than company A". The text designation is to designate the rows to be used in analysis except for the row used in category designation. For example, the text is designated by selecting the second row in Fig .2.

[0029]

Further, the text classification engine 14, by the attribute selecting means 17, selects the attributes in plural inquiry reply statements being read in from the database 13 (step 33). The attribute selection is to select a word which is important for expressing the correspondence of text and category.

[0030]

Then, the text classification engine 14 learns the rule for expressing the correspondence of text and category on the basis of the word selected by attribute by the rule learning means 18 (step 34). For example, when the category and text are designated as stated above, the rule is obtained as shown in Fig. 6.

[0031]

The rule in Fig. 6 shows that if the word "easy to use" is found in the text by reading the first line, 92.0% of the respondents assume company A as the high-tech enterprise. If the word "easy to use" is not found, next, checking if the words "future" and "private" appear at the same time, and when found, it means that 87.2% of the respondents assume company A as the high-tech enterprise. Thereafter, similarly, according the rule of if-then-else pattern, the conditional sentences are read from top to bottom. Such rule is an easy

and compact expression of the relation between the high-tech enterprise and high-tech feeling.

[0032]

Picking up other company B, when the category is designated into "company B" and "other than company B", the rule in Fig. 7 is obtained by the same procedure.

[0033]

Comparing the rule of company B in Fig. 7 with the rule of company A in Fig. 6, the high-tech feeling of the respondents assuming company A as the high-tech enterprise is mainly based on the ease of use and preference sensation, while the high-tech feeling of the respondents assuming company B as the high-tech enterprise is known to be mainly based on the efficiency. Thus, by comparing the rules, the inquiry replies can be analyzed easily.

[0034]

Finally, the text classification engine 14, by the rule output means 19, issues the knowledge of the analysis result in the rule format to the claimant (step 35).

[0035]

As an example of knowledge in rule format, herein, the stochastic decision rule is discussed, and the attribute selecting step 33 for creating it and the rule learning step 34 are more specifically described below.

[0036]

The stochastic decision list is a ranked list of stochastic rule of if-then pattern as shown in Fig. 6. Each stochastic rule has a pattern of "$c = 1 \leftarrow t$ (probability p)", where $c = 1$ is the decision of classification, t is the condition of classification, and (probability p) is the probability.

[0037]

First, the attribute selecting step 33 is explained.

[0038]

The attribute selection is to collect words closely related with the category in the given category (for example, company A and other than company A). More specifically, as shown in Fig. 4, at step 41, in each word ω appearing in the text, the difference $\Delta SC(\omega)$ between the stochastic complexity (SC) of the text set without consideration of onset of the word ω and the SC with consideration thereof is computed, and at step 42, when the difference $\Delta SC(\omega)$ is greater than the given threshold τ, the word ω is selected as an attribute.

[0039]

A practical method of computing the SC is explained. Sets of texts in the entered inquiry reply statements are expressed as

$(d_1, c_1), (d_2, c_2), \ldots, (d_m, c_m)$

where d_i denotes the i-th text, and is expressed as the row of words appearing in the i-th text, c_i denotes the value of category (label) corresponding to the i-th text, and each c_i is 1 if belonging to the given category (company A) or 0 otherwise (other than company A), and m is the number of texts.

[0040]

Further, label rows are expressed as $c^m = c_1, \ldots, c_m$, and text rows are $d^m = d_1, \ldots, d_m$. The SC of label row c^m is calculated as in formula 1.

[0041]

(1)

$$SC(c^m) = mH\left(\frac{m^+}{m}\right) + \frac{1}{2}\log\frac{m}{2\pi} + \log\pi$$

[0042]

where m^+ is the number of labels in which the value is 1 in label row c^m, and log is the natural logarithm, which is expressed in formula 2.

[0043]

(2)

$$H(z) \stackrel{def}{=} -z \log z - (1-z)\log(1-z)$$

[0044]

For example, as discussed by J. Rissanen and Fisher information and stochastic complexity (IEEE Trans. on Information Theory, 42 (1), 40-47, 1996), SC(c^m) is the shortest description length for describing the label row c^m by using the given model (herein, Bernoulli model). Suppose

[1]

$C^{m\omega}$

is a label row composed of label c_i in which word ω appears in the corresponding text d_i, where m_ω is the number of labels in

[2]

$C^{m\omega}$

Then, the value of SC in

[3]

$C^{m\omega}$

can be calculated as shown in formula 3.

[0045]

(3)

$$SC(c^{m_\omega}) = m_\omega H\left(\frac{m_\omega^+}{m_\omega}\right) + \frac{1}{2}\log\frac{m_\omega}{2\pi} + \log\pi$$

[0046]

where m_ω^+ is the number of labels of which value is 1 in

[4]

$C^{m\omega}$

On the other hand, suppose

[5]

$C^{m\neg\omega}$

is a label row composed of label c_i in which word ω does not appear in the corresponding text d_i, where m_ω is the number of labels in

[6]

$C^{m\neg\omega}$

[0047]

Then, the value of SC in

[7]

$C^{m\neg\omega}$

can be calculated as shown in formula 4.

[0048]

(4)

$$SC(c^{m_{\neg\omega}}) = m_{\neg\omega} H\left(\frac{m_{\neg\omega}^{+}}{m_{\neg\omega}}\right) + \frac{1}{2}\log\frac{m_{\neg\omega}}{2\pi} + \log\pi$$

[0049]

The difference $\Delta SC(\omega)$ between the SC without consideration of appearance of word ω and the SC with consideration thereof is calculated as shown in formula 5.

[0050]

(5)

$$\Delta SC(\omega) = \frac{1}{m}\left(SC(c^m) - (SC(c^{m_\omega}) + SC(c^{m_{\neg\omega}}))\right)$$
$$= \left[H\left(\frac{m^+}{m}\right) - \frac{m_\omega}{m} H\left(\frac{m_\omega^+}{m_\omega}\right) - \frac{m_{\neg\omega}}{m} H\left(\frac{m_{\neg\omega}^+}{m_{\neg\omega}}\right)\right]$$
$$- \left[\frac{1}{2m} \log \frac{m_\omega m_{\neg\omega} \pi}{2m}\right]$$

[0051]

The word ω large in the difference ΔSC(ω) is a word appearing very frequently or hardly in a given category. Such words are considered to be closely related with the category. Supposing τ to be a given threshold, the word ω in the relation of ΔSC(ω) > τ is selected as an attribute.

[0052]

The rule learning step 34 is explained below.

[0053]

Suppose there are n words selected of attribute, being ω_1, ..., ω_n. At step 51, first of all, sets of entered texts are expressed as follows.

$(d_1, c_1), (d_2, c_2), \ldots, (d_m, c_m)$

[0054]

where each d_i expresses a binary vector (generally, a multi-valued discrete vector)

$d_i = (\omega_{i1}, \omega_{i2}, \ldots, \omega_{in})$ (i = 1, ..., m)

where ω_{ij} is 1 when the word ω_j obtained by attribute selection appears in the i-th text, or 0 otherwise (j = 1, ..., n), c_i expresses the value (label) of the category corresponding to the i-th text, and each c_i is 1 when belonging to the specified category, and 0 otherwise, and m is the number of texts.

[0055]

At step 52, the rule of if-then-else pattern is selected, and sequentially added to the stochastic decision list A. This is called "growth." For

selection of rules, for example, the extended stochastic complexity (ESC) minimum principle is employed.

[0056]

The operation is as follows. Suppose k is a given positive integer. A set of all possible k terms (up to k link words of word ω) on the basis of the word ω by attribute selection is supposed to be T. From terms t of the set T, those not appearing in the text at all are excluded. An empty stochastic decision list A is prepared. Next, the rule of the largest decrement of ESC value is sequentially added to the stochastic decision list A.

[0057]

Herein, the ESC is computer as follows. The whole data set D is expressed as sets of data in a format of

$(d_1, c_1), (d_2, c_2), \ldots, (d_m, c_m)$

and label rows are $c^m = c_1, \ldots, c_m$. The value of the ESC of label row c_m can be approximated as in formula 6.

[0058]

(6)

$$ESC(c^m) = Loss(c^m) + \lambda\sqrt{m \log m}$$

[0059]

This is one approximate format of the original ESC proposed by K. Yamanishi in his paper (A decision-theoretic extension of stochastic complexity and its applications to learning, IEEE Trans. Inform. Theory, 44, 1424-1439, 1998).

[0060]

Herein, λ is a positive constant, Loss (c^m) is the number of errors in default classification. The default classification is to assume all labels are 0, for example. t is a term in the set T. Suppose

[0061]

[8]

c^{mt}

is a label row composed of label c_i in which term t is true in the corresponding text d_i, where m_t is the number of labels in

[9]

c^{mt}

[0062]

Suppose

[10]

$\text{Loss}(c^{mt})$

is the number of errors when classifying by term t. On the other hand,

[11]

$C^{m\neg\omega}$

is a label row composed of label c_i in which term t is false in the corresponding text d_i, where $m_{\neg t}$ is the number of labels in

[12]

$C^{m\neg\omega}$

and $\neg_t$ expresses negation of term t. Suppose

[0063]

[13]

$\text{Loss}(c^{m\neg t})$

is the number of errors when classifying by $\neg_t$. The ESC values of

[0064]

[14]

c^{mt}

and

[15]

$c^{m\neg t}$

can be calculated in formula 7 and formula 8, respectively.

[0065]

(7)

$$ESC(c^{m_t}) = Loss(c^{m_t}) + \lambda\sqrt{m_t \log m_t}$$

[0066]

(8)

$$ESC(c^{m_{\neg t}}) = Loss(c^{m_{\neg t}}) + \lambda\sqrt{m_{\neg t} \log m_{\neg t}}$$

[0067]

When classifying by term t, the decrement ΔESC(t) of the ESC value is calculated in formula 9.

[068]

(9)

$$\begin{aligned}\Delta ESC(t) &= ESC(c^m) - (ESC(c^{m_t}) + ESC(c^{m_{\neg t}})) \\ &= \left[Loss(c^m) - Loos(c^{m_t}) - Loos(c^{m_{\neg t}})\right] \\ &\quad + \left[\lambda(\sqrt{m \log m} - \sqrt{m_t \log m_t} - \sqrt{m_{\neg t} \log m_{\neg t}})\right]\end{aligned}$$

[0069]

According to the ESC minimum principle, term t is selected so that ΔESC(t) may be minimum. When such t = t* is selected, the number of data of whole data set D in which it is true is supposed to be

[16]

m_{t^*}

and of such data, the label, for example, greater in number is supposed to be c = 1, and the number of c = 1 is supposed to be

[17]

$m_{t^*}^{+}$

and the number of c = 0 is supposed to be

[18]

m_{t*}^{-}

Accordingly, the rule "c = 1 ← t* (probability)" is added to the stochastic decision list A. Herein, the probability value p is calculated, for example, as shown in formula 10 by using the method of Laplacean estimation.

[0070]

(10)

$$(m_{t\cdot}^{+} + 0.5)/(m_{t\cdot} + 1)$$

[0071]

Excluding term t* from the set T, a new set T is obtained, and excluding all data rendering term t* true from the whole data set D, a new whole data set D is obtained, and the same operation is repeated until the whole data set D becomes empty. Instead of the standard ESC used hereabove, the standard SC used in attribute selection may be used.

[0072]

At step 53, since the stochastic decision list A obtained at step 52 may excessively conform to the learning data, the rules are removed one by one from the last one of the stochastic decision list A consecutively until none should be removed from the viewpoint of the ESC minimum principle. This process is called "clipping".

[0073]

In this case, the manner of application of the ESC minimum principle is explained below. First, the value of the ESC corresponding to the stochastic decision list A of label row c^m is defined as the sum of ESC values corresponding to all terms t in the stochastic decision list A as shown in formula 11.

[0074]

(11)

$$ESC(c^m \mid A) = \sum_t ESC(c^{m_t})$$

[0075]

However,

[19]

ESC(c^{mt})

is calculated as in formula 7. Next, the whole ESC value of label row c^m and stochastic decision list A are defined as in formula 12.

[0076]

(12)

$$ESC(c^m : A) = ESC(c^m \mid A) + \lambda' L(A)$$
$$= \sum_t Loss(c^{m_t}) + \lambda \sum_t \sqrt{m_t \log m_t} + \lambda' L(A)$$

[0077]

where λ' is a positive constant, and L(A) is a code length for encoding the stochastic decision list A. Specifically, it is calculated as L(A) = logT + log(T-1) + ... + logT (T-i+1), where T is the number of possible terms t, and i is the number of rules in the stochastic decision list A.

[0078]

Suppose A expresses the stochastic decision list before clipping, and A' is the stochastic decision list after clipping.

ESC (c^m|A) ≦ ESC (c^m|A')

[0079]

That is, as far as

ESC (c^m|A') - ESC (c^m|A) ≧ λ' (L(A) - L(A'))

is established, the clipping procedure continues, and when this condition is no longer satisfied or there is no rule left to be clipped, the stochastic decision list A obtained at this moment is delivered. Thus, the stochastic

decision list A small in the ESC on the whole is issued.

[0080]

In the inquiry reply analysis system of the first embodiment, rules of analysis results can be automatically extracted from the inquiry reply statements including free reply description in natural language collected through the network 12.

[0081]

In the inquiry reply analysis system of the first embodiment, as the text classification engine 14, by using the engine proposed by Li and Yamanishi (H. Li and K. Yamanishi: Text Classification Using ESC-based Stochastic Decision Lists, Proceedings of 1999 International Conference on Information & Knowledge Management, pp. 122-130, 1999), by the computation amount of O ($n^k m$), rules can be extracted from the inquiry reply statements at high speed, where n is the number of words of attribute selection from the inquiry reply statements, m is the number of inquiry reply statements, and k is the maximum number of words included in the link words relating to one condition. Hence, efficient automatic analysis of inquiry reply statements is possible. The obtained rules can express the inquiry reply statements belonging to a specific category in compact and easy format of if-then-else pattern.

[0082]

The inquiry reply analysis system of the first embodiment can be applied, for example, in the following business. Receiving orders for enterprise image survey or inquiry about specific merchandise or service from clients, the inquiry of the items as shown in Fig. 2 is operated on the network 12, and the inquiry reply statements including free reply description in natural language collected online through the network 12 are accumulated in the database 13, and inquiry reply statements are called therefrom, and the rules obtained by using the text classification engine 14 are sold to the

clients as the analysis results.

[0083]

(2) Second Embodiment

Fig. 8 is a block diagram showing a configuration of an inquiry reply analysis system according to a second embodiment of the invention. The inquiry reply analysis system of the embodiment mainly comprises inquiry reply input means 81, a database 82, and a text classification engine 83.

[0084]

The inquiry reply input means 81 is directly connected to the database 82 without connecting through network.

[0085]

The database 82 accumulates inquiry reply statements from plural inquiry respondents.

[0086]

The text classification engine 83 is exactly the same as the text classification engine 14 in the inquiry reply analysis system of the first embodiment shown in Fig. 1. Therefore, the corresponding parts are identified with same reference numerals, and their detailed description is omitted.

[0087]

The operation of the inquiry reply analysis system of the second embodiment having such configuration is explained below.

[0088]

The inquiry reply input means 81 is directly connected to the database 82 without connecting through network, and receives inquiry reply statements including free reply description in natural language.

[0089]

The database 82 accumulates inquiry reply statements from plural inquiry respondents.

[0090]

The text classification engine 83 reads out plural inquiry reply statements from the database 82, extracts the rules for classifying the inquiry reply statements, and issues the rules of analysis result to the claimant. The detail of the operation of the text classification engine 83 is same as that of the text classification engine 14 of the inquiry reply analysis system of the first embodiment, and the detailed description is omitted.

[0091]

The inquiry reply analysis system of the second embodiment can be applied, for example, in the following business. Receiving orders for enterprise image survey or inquiry about specific merchandise or service from clients, the inquiry of the items as shown in Fig. 2 is operated, and the inquiry reply statements including free reply description in natural language are collected at once, and accumulated in the database 82, and inquiry reply statements are called therefrom, and the analysis results obtained by using the text classification engine 83 are sold to the clients.

[0092]

(3) Third Embodiment

Fig. 9 is a block diagram showing a configuration of an inquiry reply analysis system according to a third embodiment of the invention. The inquiry reply analysis system of the embodiment mainly comprises respondent computers 911 to 91N, a network 92, a database 93, a text classification engine 94, and a claimant computer 95.

[0093]

The respondent computers 911 to 91N are computers, portable information terminals, cellular phones, and other devices having transmission function of message, mail and the like, which are connected to the network 92.

[0094]

The network 92 includes various networks, whether wired or wireless, such as public networks, exclusive networks, or LAN.

[0095]

The database 93 is connected to the network 92, and inquiry reply statements from plural respondents transmitted from the respondent computers 911 to 91N through the network 92 are accumulated herein.

[0096]

The text classification engine 94 is same as the text classification engine 14 in the inquiry reply analysis system of the first embodiment shown in Fig. 1, except that the rule output means 19 can transmit the knowledge of the rule format as a result of analysis through the network 92. Therefore, same reference numerals are given to the corresponding parts and detailed description is omitted.

[0097]

The claimant computer 95 requests knowledge of rule format as a result of analysis to the text classification engine 94 through the network 92, and receives the knowledge of rule format of analysis result from the text classification engine 94 through the network 92.

[0098]

The operation of the inquiry reply analysis system of the third embodiment having such configuration is explained below.

[0099]

The inquiry respondents send inquiry reply statements including free reply description in natural language from respondent computers 911 to 91N through the network 92. Suppose the number of respondents to be N.

[0100]

The database 93 is connected to the network 92, and accumulates inquiry reply statements from plural inquiry respondents.

[0101]

The text classification engine 94 reads out plural inquiry reply statements from the database 93, an extracts the knowledge of rule format for classifying the inquiry reply statements. The text classification engine 94 is connected to the network 92, and distributes the knowledge of rule format of analysis result through the network 92 depending on the request from the claimant computer 95. The detail of operation of the text classification engine 94 is same as that of the text classification engine 14 of the inquiry reply analysis system of the first embodiment, except that the knowledge of rule format of analysis result is distributed through the network 92, and the description of detail is omitted.

[0102]

The inquiry reply analysis system of the third embodiment can be applied, for example, in the following business. Receiving orders for enterprise image survey or inquiry about specific merchandise or service from clients, the inquiry of the items as shown in Fig. 2 is operated on the network 92, and the inquiry reply statements including free reply description in natural language collected online through the network 92 are accumulated in the database 93, and inquiry reply statements are called therefrom, and the analysis results obtained by using the text classification engine 94 are distributed through the network 92 to the clients when requested.

[0103]

(4) Fourth Embodiment

Fig. 10 is a block diagram showing a configuration of an inquiry reply analysis system according to a fourth embodiment of the invention. The inquiry reply analysis system of the embodiment is similar to the inquiry reply analysis system of the first embodiment shown in Fig. 1, except that a recording medium 102 recording a text classification engine program is incorporated in a computer 101 connected to the database 13, and the other composition is same as that of the inquiry reply analysis system of the first

embodiment, and corresponding parts are identified with same reference numerals and detailed description is omitted.

[0104]

In the inquiry reply analysis system of the fourth embodiment having such configuration, the text classification engine program is read into the computer 101 from the recording medium 102, and controls the operation of the computer 101 as the text classification engine 14 including the morpheme analysis means 15, category-text designating means 16, attribute selecting means 17, rule learning means 18, and rule output means 19. The detail of operation of the text classification engine 14 on the computer 101 is exactly same as in the case of the inquiry reply analysis system of the first embodiment, and detailed description is omitted.

[0105]

(5) Fifth Embodiment

Fig. 11 is a block diagram showing a configuration of an inquiry reply analysis system according to a fifth embodiment of the invention. The inquiry reply analysis system of the embodiment is similar to the inquiry reply analysis system of the second embodiment shown in Fig. 8, except that a recording medium 112 recording a text classification engine program is incorporated in a computer 111 connected to the database 82, and the other composition is same as that of the inquiry reply analysis system of the second embodiment, and corresponding parts are identified with same reference numerals and detailed description is omitted.

[0106]

In the inquiry reply analysis system of the fifth embodiment having such configuration, the text classification engine program is read into the computer 111 from the recording medium 112, and controls the operation of the computer 111 as the text classification engine 83 including the morpheme analysis means 15, category-text designating means 16, attribute

selecting means 17, rule learning means 18, and rule output means 19. The detail of operation of the text classification engine 83 on the computer 111 is exactly same as in the case of the inquiry reply analysis system of the second embodiment, and detailed description is omitted.

[0107]

(6) Sixth Embodiment

Fig. 12 is a block diagram showing a configuration of an inquiry reply analysis system according to a sixth embodiment of the invention. The inquiry reply analysis system of the embodiment is similar to the inquiry reply analysis system of the third embodiment shown in Fig. 9, except that a recording medium 122 recording a text classification engine program is incorporated in a computer 121 connected to the database 93, and the other composition is same as that of the inquiry reply analysis system of the third embodiment, and corresponding parts are identified with same reference numerals and detailed description is omitted.

[0108]

In the inquiry reply analysis system of the sixth embodiment having such configuration, the text classification engine program is read into the computer 121 from the recording medium 122, and controls the operation of the computer 121 as the text classification engine 94 including the morpheme analysis means 15, category-text designating means 16, attribute selecting means 17, rule learning means 18, and rule output means 19. The detail of operation of the text classification engine 94 on the computer 121 is exactly same as in the case of the inquiry reply analysis system of the third embodiment, and detailed description is omitted.

[0109]

[Effects of the Invention]

[0110]

According to a first effect of the invention, by receiving orders for

enterprise image survey or inquiry about specific merchandise or service from clients, the inquiry is operated on the network, and the inquiry reply statements including free reply description in natural language collected online through the network are accumulated in the database, and inquiry reply statements are called therefrom, and the rules obtained by using the text classification engine are sold to the clients as the analysis results.

[0111]

According to a second effect of the invention, by receiving orders for enterprise image survey or inquiry about specific merchandise or service from clients, the inquiry is operated, and the inquiry reply statements including free reply description in natural language are collected at once, and accumulated in the database, and inquiry reply statements are called therefrom, and the analysis results obtained by using the text classification engine are sold to the clients.

[0112]

According to a third effect of the invention, by receiving orders for enterprise image survey or inquiry about specific merchandise or service from clients, the inquiry is operated on the network, and the inquiry reply statements including free reply description in natural language collected online through the network are accumulated in the database, and inquiry reply statements are called therefrom, and the analysis results obtained by using the text classification engine are distributed through the network to the clients when requested.

[Brief Description of the Drawings]

[Fig. 1]

It is a block diagram showing a configuration of an inquiry reply analysis system according to a first embodiment of the present invention.

[Fig. 2]

It is a diagram showing an example of inquiry reply statements

accumulated in a database in Fig. 1.

[Fig. 3]

It is a flowchart showing processing in a text classification engine in Fig. 1.

[Fig. 4]

It is a flowchart showing a more specific processing of attribute selecting step in Fig. 3.

[Fig. 5]

It is a flowchart showing a more specific processing of rule learning step in Fig. 3.

[Fig. 6]

It is a diagram showing an example of rule format knowledge (stochastic decision list) as a result of analysis by the text classification engine in Fig. 1.

[Fig. 7]

It is a diagram showing other example of rule format knowledge (stochastic decision list) as a result of analysis by the text classification engine in Fig. 1.

[Fig. 8]

It is a block diagram showing a configuration of an inquiry reply analysis system according to a second embodiment of the present invention.

[Fig. 9]

It is a block diagram showing a configuration of an inquiry reply analysis system according to a third embodiment of the present invention.

[Fig. 10]

It is a block diagram showing a configuration of an inquiry reply analysis system according to a fourth embodiment of the present invention.

[Fig. 11]

It is a block diagram showing a configuration of an inquiry reply analysis system according to a fifth embodiment of the present invention.

[Fig. 12]

It is a block diagram showing a configuration of an inquiry reply analysis system according to a sixth embodiment of the invention.

[Explanation of Reference Numerals]

12 NETWORK
13 DATABASE
14 TEXT CLASSIFICATION ENGINE
15 MORPHEME ANALYSIS MEANS
16 TEXT-CATEGORY DESIGNATING MEANS
17 ATTRIBUTE SELECTING MEANS
18 RULE LEARNING MEANS
19 RULE OUTPUT MEANS
31 MORPHEME ANALYSIS STEP
32 TEXT AND CATEGORY DESIGNATING STEP
33 ATTRIBUTE SELECTING STEP
34 RULE LEARNING STEP
35 RULE OUTPUT STEP
41 $\Delta SC(\omega)$ COMPUTING STEP
42 ATTRIBUTE SELECTING STEP
51 DATA FORMING STEP
52 GROWTH PROCESSING STEP
53 CLIP PROCESSING STEP
81 INQUIRY REPLY INPUT MEANS
82 DATABASE
83 TEXT CLASSIFICATION ENGINE
92 NETWORK
93 DATABASE
94 TEXT CLASSIFICATION ENGINE
95 CLAIMANT COMPUTER

101, 111, 121 COMPUTER

102, 112, 122 RECORDING MEDIUM

111 TO 11N RESPONDENT COMPUTER

911 TO 91N RESPONDENT COMPUTER

5

[Name of the Document] Abstract

[Abstract]

[Problem] To analyze inquiry replies automatically from inquiry reply statements including free reply description in natural language collected through a network, by using a text classification engine, and to issue the analysis results of the claimant as the knowledge of rule formats.

[Solving Means] Inquiry respondents send inquiry reply statements from respondent computers 111 to 11N. Inquiry reply statements are accumulated in a database 13 through a network 12. A text classification engine 14 reads out accumulated inquiry reply statements from the database 13, and learns the rules for classifying the inquiry reply statements, and issues to the claimant.

[Selected Drawing] Fig. 1

[Name of Document] Drawings

FIG.1

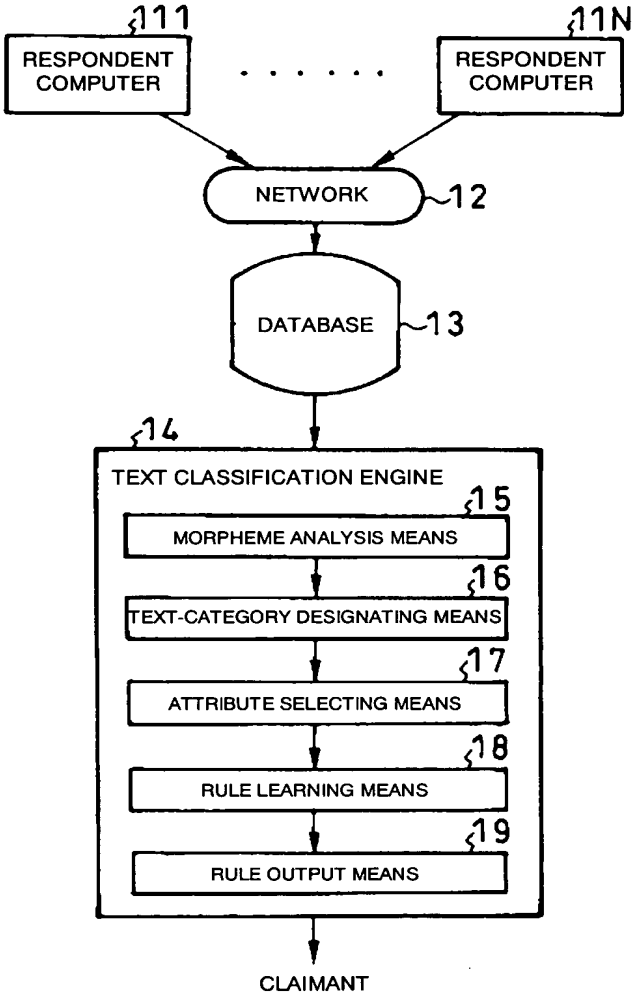


FIG.2

INQUIRY RESPONDENT	WHICH DO YOU ASSUME AS HIGH-TECH ENTERPRISE?	WHAT'S HIGH-TECH FOR YOU?	WHAT DO YOU ASSUME AS HIGH-TECH PRODUCT?	....
1	COMPANY A	ADVANCED AND FUTURISTIC MACHINE	ROBOT	.....
2	COMPANY C	EASY AND FRIENDLY MACHINE	CELL PHONE	.....
3	COMPANY A	HIGH SPEED AND HIGH PERFORMANCE MACHINE	PERSONAL COMPUTER	.....

FIG.3

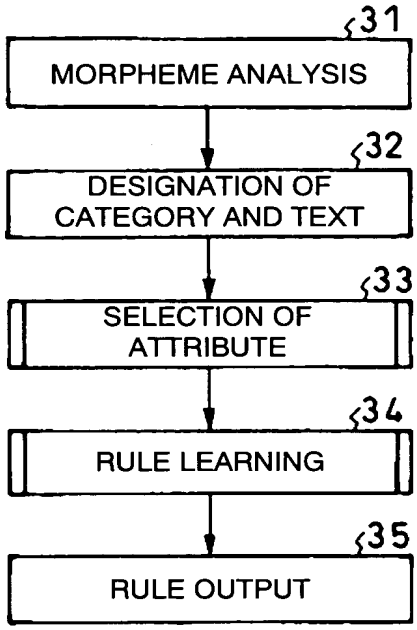


FIG.4

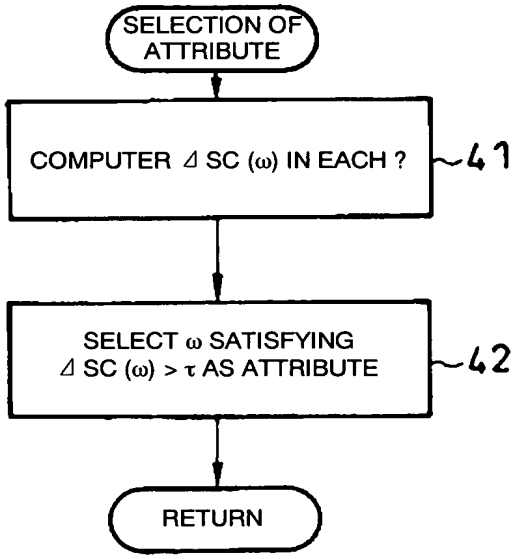


FIG.5

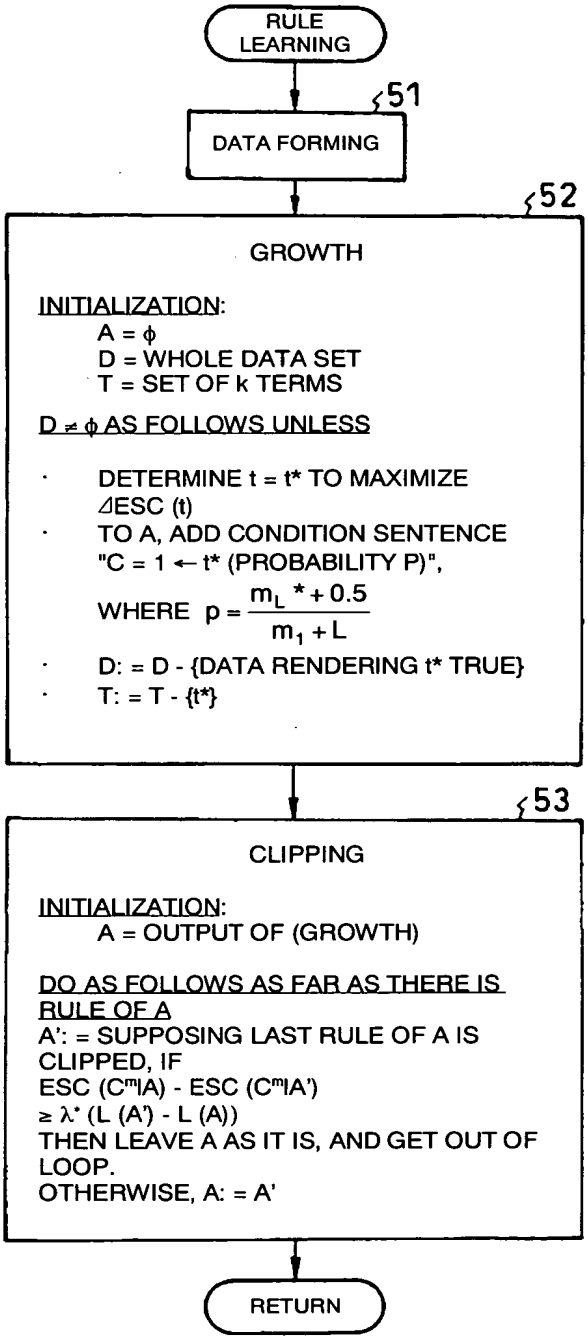


FIG.6

COMPANY A	←	EASY TO USE	[92.0%]
COMPANY A	←	FUTURE & PRIVATE	[87.2%]
COMPANY A	←	FATIGUE & RELIEF	[78.0%]
COMPANY A	←	EASY	[65.8%]
COMPANY A	←	PLEASANT	[56.2%]
OTHER THAN COMPANY A	←	OR ELSE	[79.4%]

FIG.7

COMPANY B	←	QUICK	[82.0%]
COMPANY B	←	MACHINE & EFFICIENCY	[77.8%]
COMPANY B	←	MACHINE & MANIPULATION	[76.0%]
COMPANY B	←	CLEVER	[60.8%]
COMPANY B	←	EXCELLENT	[60.2%]
OTHER THAN COMPANY B	←	OR ELSE	[76.4%]

FIG.8

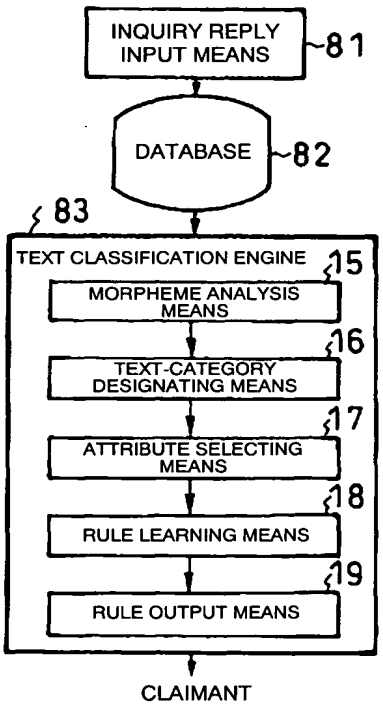


FIG.9

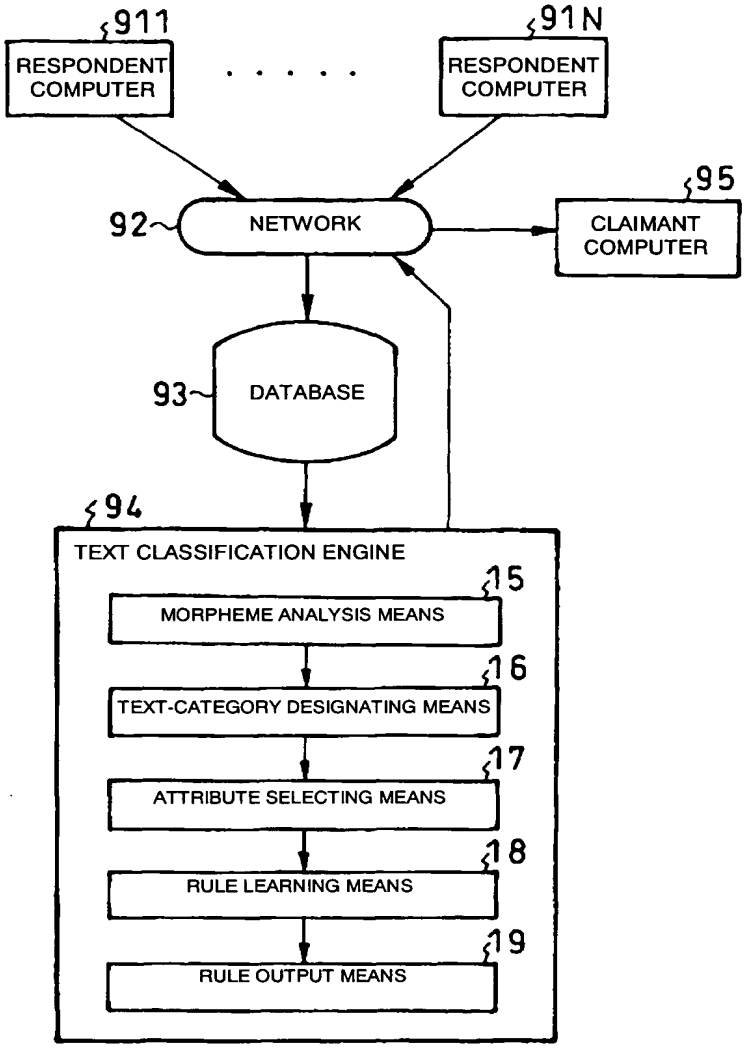


FIG.10

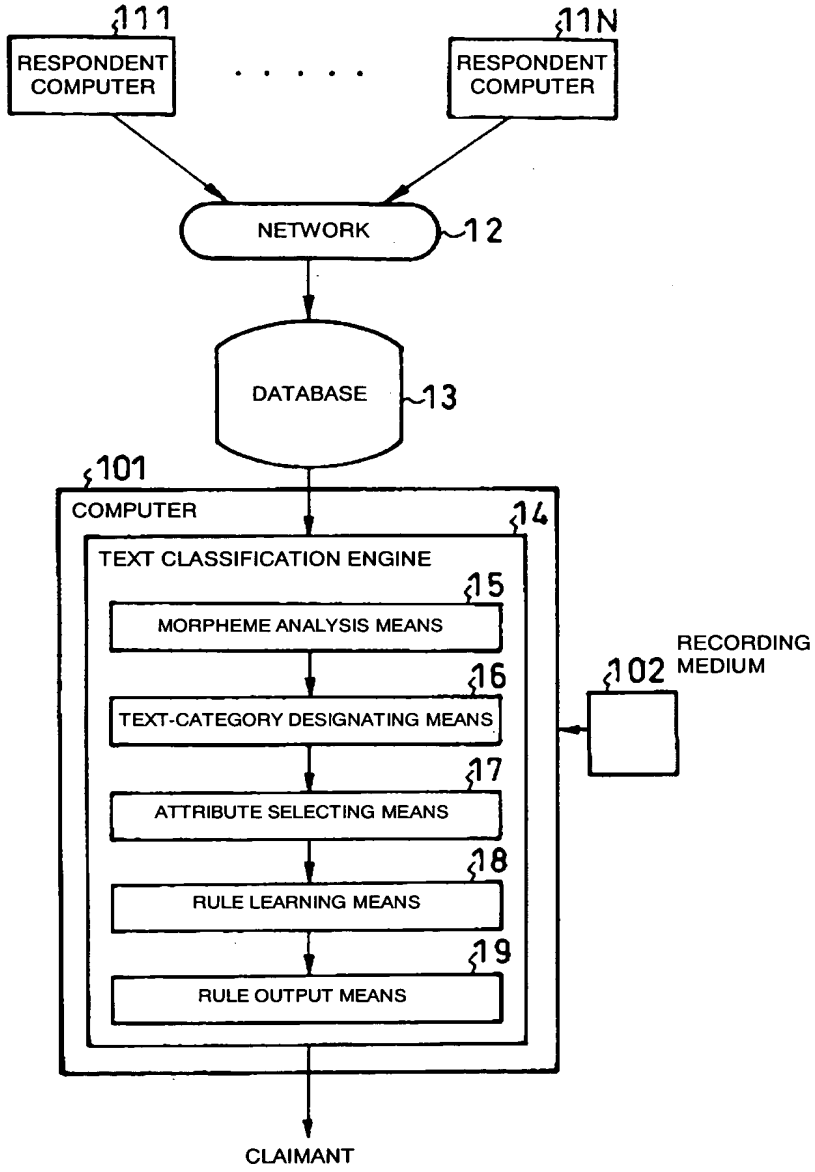


FIG.11

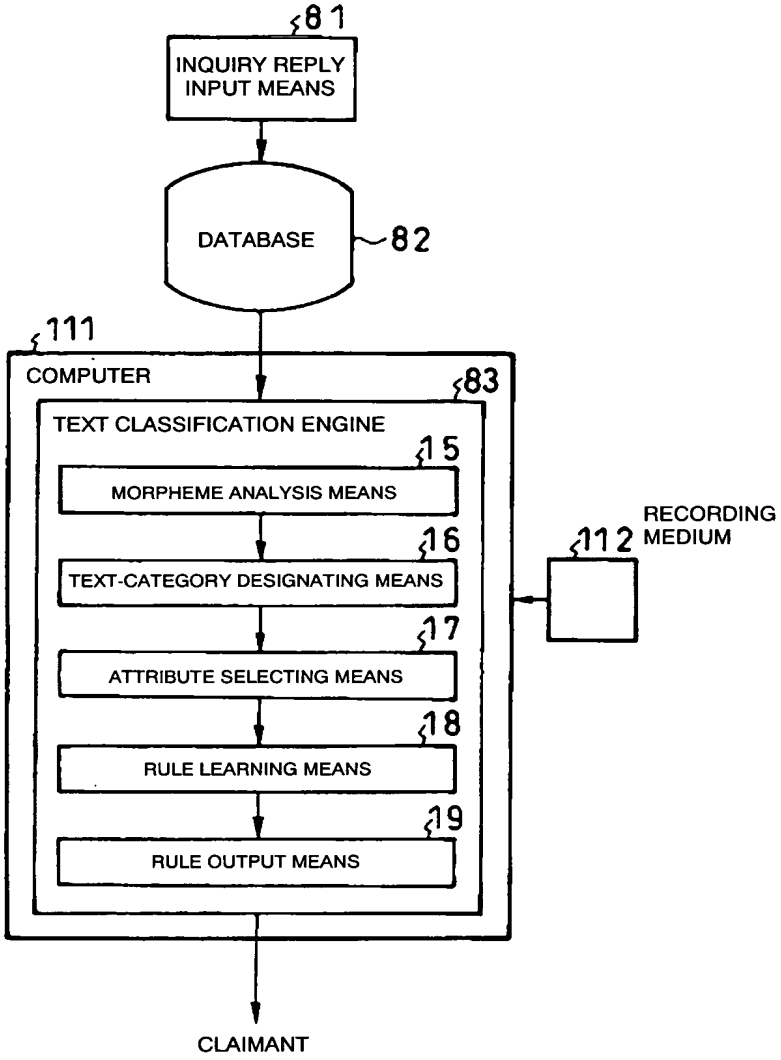


FIG.12